

DOCUMENT RESUME

ED 414 798

HE 030 691

AUTHOR Chang, Te-Sheng
TITLE Student Ratings of Instruction at Teachers College in Taiwan.
PUB DATE 1997-09-19
NOTE 17p.
PUB TYPE Reports - Research (143)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Academic Achievement; *Class Size; College Faculty; Course Content; Foreign Countries; Higher Education; Lesson Plans; Predictor Variables; *Sex Differences; Student Evaluation; *Student Evaluation of Teacher Performance; Teaching Methods
IDENTIFIERS National Hualien Teachers College (Taiwan)

ABSTRACT

This study examined the effects of course level, class size, and instructor gender on student ratings of teacher performance at the National Hualien Teachers College in Taiwan. A total of 9,843 usable evaluation forms were completed for 440 courses (96 freshmen-level, 142 sophomore-level, 119 junior-level, and 83 senior-level) taught by 132 instructors. The evaluation form contained 13 questions that focused on four dimensions of teacher performance: preparation/planning, material/content, method/skill, and assignment/examination. The results indicated that as the level of the course increased, the mean for each evaluation dimension increased. Class size had no significant effect on student ratings of instruction, though the highest mean evaluations came from classes with 10 or fewer students. Instructor gender had no statistically significant effect on student ratings of instruction, though male instructors were rated significantly higher on the material/content dimension. (Contains 19 references.) (MDM)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED 414 798

Student Ratings of Instruction at Teachers College in Taiwan

Te-Sheng Chang

National Hualien Teachers College, Hualien, Taiwan, ROC

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

* Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
IN OTHER THAN PAPER COPY
HAS BEEN GRANTED BY

Te-Sheng Chang

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

Correspondence concerning this article should be addressed to Te-Sheng Chang,
National Hualien Teachers College, Hualien, Taiwan, 970 R.O.C. or contact by
E-mail: achang@sparc2.nhlte.edu.tw

Abstract

Student Ratings of Instruction have been implemented at National Hualien Teachers College since the fall semester of the 1996-1997 academic year. The purpose of this study was to investigate the effect of course level, class size, and instructor gender on student ratings of teaching performance in this school. The sample consisted of 440 courses rating 132 instructors with 96 freshman classes, 142 sophomore classes, 119 junior classes, and 83 senior classes. The Student Ratings of Instruction (SRI) form was used to measure students' perceptions of faculty performance. The rating form was composed of 13 questions rated on a 5-point Likert scale ranging from strongly agree to strongly disagree. It yielded four dimensions: Preparation/Planning, Material/Content, Method/Skill, and Assignment/Examination. One-way analysis of variance was applied to test the effect of course level and class size; independent t test was applied to test the impact of instructor gender. The results of this study indicated that (a) as the level of courses increased the mean for each evaluation dimension increased; (b) class size had no significant effect on student ratings of instruction, and (c) there was no statistical difference between courses taught by female instructors and courses taught by male instructors on all evaluation dimensions except that male instructors were significantly rated higher in Material/Content. The technical and practical ramifications of these findings were discussed.

Key words: Student Ratings, Course Evaluation, Faculty Evaluation, College Teaching Teachers College

Student Ratings of Instruction at Teachers College in Taiwan

Student ratings have been widely used as a measure of teaching effectiveness in universities and colleges in western countries. They can be useful for several purposes, such as: providing data for improving teaching and evaluating faculty performance; aiding student choice of courses and professors, and stimulating students to think about their education (Schlenker & Mckinnon, 1994). Unlike most universities in western countries which have a long history in applying student ratings to teacher effectiveness, student ratings of instruction have just been employed at National Hualien Teachers College in Taiwan since the fall semester of the 1996-1997 academic year. The school evaluation committee has assigned the weights of 40%, 40%, and 20% to its faculty for research, teaching, and service, respectively. The results of student ratings are assigned as a major component of faculty teaching performance.

The weakness of student ratings of instruction, according to literature and the researcher's experience, is that they are impacted by some course characteristics, such as, class size, course level, and instructor gender, over which teachers have little control. The effect of these course characteristics on student ratings of instruction is important information in determining how student rating should be reported.

Many of the studies showed that the higher the course level the higher the average student rating of the teacher and the course. Basow and Silberg (1987) reported that students gave increasingly better evaluation to their instructors as they moved from freshman to senior status. A study by Schlenker and Mckinnon (1994) presented similar positive relationships: as course level increased, student satisfaction with instructors' performance also increased. It should be noted, as Feldman (1978) pointed out, "although the positive association between course level and ratings is

clear and relatively consistent across various rating items in almost all of these studies, it does tend to be quite weak in strength" (p. 216).

A positive association between class ratings and course level is not universally found (Feldman, 1978). There are studies in which course level is essentially unrelated to class ratings (Doyle & Whitely, 1974). Results in the study by Marsh, Overall, and Thomas (1976) were basically inconsistent across rating items, with indications in their study that the higher the course level the lower the ratings on certain items. However, the finding by McKeachie (1986) was that course level had little impact on student evaluation of teaching performance.

Many studies of class size on student ratings have produced conflicting results. Lovell and Haner (1955) found the positive relationship between class size and student ratings and Riley, Ryan, and Lifschitz (1950) found the negative relationship. Jiobu and Pollis (1971) stressed that little if any relationship between class size and student ratings. Centra (1979) concluded that classes of fifteen and fewer were more effective in producing student learning and that teachers were rated higher by students. Schlenker and McKinnon (1994) reported that the effect of class size on student ratings varied for different evaluation dimensions. However they concluded that the optimal class size may be 16-30 students. Marsh, Overall, and Thomas (1976) found that there were small but statistically significant relationships between course enrollment and 10 factors of effective instruction. Linear correlations with course enrollment accounted for 13% of the variance in an "effectiveness of student-teacher interaction" factor, but less than 5% of the variance in any of nine other factors, including overall rating of the course and instructor. In another study, Marsh, Overall, and Kesler (1979) found that the form of the class-size effect was clearly nonlinear.

Research has investigated the effect of faculty gender on student ratings of teaching performance in colleges since the 1960s. Downie (1952) found that female faculty compared to male faculty received significantly higher ratings for the extent to which they brought new books and authors into the classroom. Elmore and LaPointe (1974) found that female faculty received significantly higher ratings on returning homework and tests promptly, while male faculty received significantly higher ratings on giving clear speech. Kaschak (1978) found that female students rated female and male professors as equally effective, concerned, liable and excellent. In contrast, male student showed a consistent bias in favor of male professors. In a second study, Kaschak (1981) found that male professors were still described as more powerful and effective than female professors. Lombardo and Tocci (1979) found that male instructors were perceived as more competent than female instructors by male students, and whereas female students showed no sex bias. Although these studies have found a gender difference on student ratings, some have found relatively few or no differences in the evaluations of male and female professors on the basis of gender alone (Basow & Distenfeld, 1985; Bennet, 1982; Elmore & Lapointe, 1975).

Despite the controversy generated over time, numerous research studies have continued focusing on the field of student ratings of teacher performance. The results from the research literature reviewed in this study suggested that course level, class size, and instructor gender play a role in student ratings of college teachers. The purpose of this study was to disseminate recent research findings which were based on students' ratings of faculty performance at National Hualien Teachers College. The influence of three course characteristics: course level, class size, and instructor gender, on student ratings of teaching performance were investigated in the study.

Method

The data in this study were adopted from the Student Ratings of Instruction Program at National Hualien Teachers College in Taiwan. Because 77 classes of questionnaires were eliminated due to incomplete responses, the final sample consisted of 440 courses rating 132 instructors at National Hualien Teachers College, with 96 (22%) freshman classes, 142 (32%) sophomore classes, 119 (27%) junior classes, and 83 (19%) senior classes. It was possible that one instructor was rated by several courses and that one student contributed several ratings to the database. Given the sample size, it was expected that the effects of these repeated observations would be negligible. Graduate courses and courses with enrollment less than five were not included in this study. Seven Departments included in the study were: Early Childhood Education, Elementary Education, Fine Art Education, Language and Literature Education, Music Education, Mathematics and Science Education, and Social Studies Education. Both full-time and part-time instructors were evaluated.

The Student Ratings of Instruction (SRI) form used to measure students' perceptions of faculty performance was developed by the Academic Office of National Hualien Teachers College. It was composed of 13 questions rated on a 5-point Likert scale ranging from strongly agree to strongly disagree. It yielded four dimensions: Preparation/Planning, Material/Content, Method/Skill, and Assignment/Examination. The α coefficients of internal consistency reliability from pilot study were .71, .86, .84, and .82 for Preparation/Planning, Material/Content, Method/Skill, and Assignment/Examination, respectively. These coefficients confirmed that the questionnaire was a reliable instrument.

To maximize respondent participation, the evaluation forms were administered by the Vice-President of the Academic Office and her assistants during the last two

weeks of the semester. The instructor of record was absent from the classroom during the evaluation process. Anonymity of responses was emphasized. When the forms were completed they were returned to the administrator, placed in an enveloped and sealed, prior to returning them to the Academic Administrative Office. This procedure yielded 9843 useable evaluation forms which were entered into the SPSS (Statistical Package for Social Science) analysis system, a statistical software package.

This study attempted to determine the effect of three variables on student ratings of instruction. The three independent variables were course level, class size, and instructor gender. The course level, according to the Academic Administrative Office, were gauged by 1000, 2000, 3000, and 4000. Class size was measured by the number of evaluations returned per class. The data were categorized by size of class and arranged into four groups: (1) 10 or less, (2) 11 to 20, (3) 21 to 30, (4) 31 or more. Instructor gender referred to all the male instructors and female instructors who taught at National Hualien Teachers College in the fall semester of the 1996-1997 academic year. Each of these three variables was investigated in relation to the evaluation dimensions of teacher performance.

Results

The results of the data analysis exhibited in Tables 1-3 present means and standard deviations for each category of class size, course level, and instructor gender and each of the four evaluation dimensions. Additionally, the *F*-values for class size and course level and *t*-values for instructor gender were displayed, enabling an assessment of difference between means of the categories of the independent variable for each evaluation dimension.

Table 1 showed that the differences in means among the four course levels for all evaluation dimensions. In Preparation/Planning, the test was significant ($F=7.919, p$

< .0001) and the means were 3.864, 3.951, 3.967, and 4.146 for the 1000 level, 2000 level, 3000 level, and 4000 level courses, respectively. The post hoc Scheffé procedure indicated that the 4000 level courses significantly received the highest mean ratings, followed in order by 3000 level, 2000 level, and 1000 level courses. The overall test for Material/Content was statistically significant ($F = 10.398$, $p < .0001$). The 4000 level courses had the highest mean (4.287) and the 1000 level courses had the lowest mean (3.967). The results showed that as the course level increased the mean for Method/Skills increased ($F = 12.357$, $p < .0001$). As the course level advanced the mean level of satisfaction with Assignment/Examination increased ($F = 6.266$, $p < .0001$).

Table 2 displayed that there were no significant differences among the four class sizes in each of evaluation dimensions. The p value of overall F tests for Preparation/Planning, Material/Content, Method/Skills, and Assignment/Examination was .573, .888, 0.209, and .079, respectively. Nevertheless, among four different class sizes, the class with ten or less students had the highest means of Material/Content (4.123), Method/Skill (3.946), and Assignment/Examination (4.000). In Preparation/Planning, the class with ten or less students received the second highest mean (3.994) next to the one from the 31+ class (4.002).

Table 3 showed that the male instructor (4.133) was rated significantly higher than the female instructor (4.049) in Material/Content ($t = 2.135$, $p < .05$). In the rest of dimensions, instructor gender did not reach an acceptable level of significance and there was little difference between male instructor and female instructor for Preparation/Planning ($t = .978$, $p = .329$), Method/Skills ($t = 1.723$, $p = .086$), and Assignment/Examination ($t = 1.215$, $p = .225$). However, the male instructor was rated higher than the female instructor in Preparation/Planning, Material/Content, and

Assignment/Examination. The female instructor received higher rating scores than the male instructor in Method/Skills.

Conclusions

This study provided a baseline for exploring answers to the questions related to how course level, class size, and instructor gender influenced the student ratings of instruction at National Hualien Teaches College. Interestingly, course level was consistently the variable in the analysis that displays the great affect on the dimensions of student evaluation. The analyses of responses indicated that ratings were higher for 4000 level course than other level courses. As the course level consistently increased, the rating scores increased in each of four evaluation dimensions. As other research has found (Basow & Silberg, 1987; Schlenker and Mckinnon, 1994), students gave increasingly better evaluation to their instructors as they moved from freshman courses to senior courses. This finding tended to run counter to McKeachie's (1986) research concluding that course level had no impact upon student evaluation of teaching performance.

Why did the ordering of means cross the course levels? One of possible explanations from Schlenker and Mckinnon (1994) was as follows:

"This is understandable, in that, 1000 level courses are introductory, larger in size, may be core requirements and should provide opportunities for students to develop substantive areas of learning. All of these factors may contribute to lower evaluation scores." Also, in upper level courses, students are older, more totally immersed in their discipline, have been exposed to a variety of teaching styles, more fully appreciate the complexity of the classroom experience, and are more serious about accomplishing their educational goals. (p. 17)

Findings indicated that the class size had no impact on student ratings of instruction across all four evaluation dimensions although the highest mean evaluation came from the class with ten or less students in Material/Content, Method/Skill and Assignment/Examination. In terms of statistical significance, this study agreed with the results found by Jiobu and Pollis (1971); there is no significant relationship between class size and student ratings. However, in both Assignment/Examination and Method/Skill, it might tend to bear positively on previous research findings that indicated as class enrollment increased, student ratings of instruction decreased (Centra, 1979).

The means for each class size level decreased as class size increased in both Assignment/Examination and Method/Skill. Conceivably, larger classes provided less opportunities for individualized attention, mentoring, peer interaction and student initiated discussion. On the other hand, classes with less than ten students may allow for individualized instruction and the establishment of rapport. The smallest and the largest classes were rated more favorably than the other two level classes in Preparation/Planning and Material/Content. This could support the research conducted by Marsh, Overall, and Kesler (1979): The form of the class-size effect was clearly nonlinear. The magnitude of the class-size effect varied with different dimensions of student ratings. Since the effect of class size on student ratings was not statistically significant, further study should be conducted to confirm the finding before making any further conclusion.

The male instructor was rated significantly higher than the female instructor in Material/Content only in this study. This finding was similar to the studies by Kaschak (1978) and Lombardo and Tocci (1979). One of possible explanations was

that college teaching has been viewed as a male occupation. Freeman (1994) pointed out that less favorable ratings of woman were likely to occur when woman entered occupations seen as not fitting their stereotype. Except in Material/Content, no significant effects involving instructor gender were found in the other three evaluation dimensions. This was similar to the findings by Basow, Distenfeld (1985), Bennet (1982), Elmore, and Lapointe (1975). Although a case could still be made that students perceive female faculty less favorably than they do male faculty in Material/Content, it does appear that students did not evaluate an instructor on the basis of gender only.

This paper suggests that the report of student ratings should be separated for the different course level. The effect of class size on student ratings of instruction needs to be investigated to confirm the findings. Although instructor gender creates a trivial effect on student ratings in this study, the interaction between other variables and instructor gender should be examined, for example student gender, or teacher rank. It must be also acknowledged that other factors impinge upon student ratings of instruction, such as, student motivation, ability, achievement, curriculum area, and course format.

Student ratings of instruction are the one of most popular ways to evaluate the quality of college teaching, some certain cautions need to be considered. First, their use can negatively impact education if they generate anxiety or conflict on the part of faculty or students. Second, student ratings should only be used in conjunction with multiple measures of faculty performance. Third, when student evaluation of instruction is used to compare between teachers in different courses, the results should be interpreted across similar course levels, and where appropriate, instructor gender.

References

- Basow, S. A., & Distenfeld, M. S. (1985). Teacher expressiveness: More important for males than females. Journal of Educational Psychology, 77(1), 308-314.
- Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? Journal of Educational Psychology, 79(2), 308-314.
- Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluations. Journal of Educational Psychology, 74(2), 170-179.
- Centra, J.A (1979). Determining faculty effectiveness. San Francisco: Jossey-Bass.
- Dolye, K. O., & Whitely, S. E. (1974). Student ratings as criteria for effective teaching. American Educational Research Journal, 11, 259-274.
- Downie, N. W. (1952). Student evaluation of faculty. Journal of Higher Education, 23, 495-496.
- Elmore, P. B., & LaPointe, K. A. (1974). Effects of teacher gender and student gender on the evaluation of college instructors. Journal of Educational Psychology, 66(2), 386-389.
- Elmore, P.B., & LaPointe, K. A. (1975). Effects of teacher gender, student gender, and teacher warmth on the evaluation of college instructors. Journal of Educational Psychology, 67(2), 368-374.
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. Research in Higher Education, 9, 199-242.

Freeman, H. (1994). Student Evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. Journal of Educational Psychology, 86(4), 627-630.

Kaschak, E. (1978). Gender bias in student evaluation of college professors. Psychology of Women Quarterly, 3, 235-243.

Jiobu, R. M., & Pollis, C. A. (1971). Student evaluations of courses and instructors. The American Sociologist, 6, 317-321.

Lombardo, J., & Tocci, M. (1979). Attribution of positive and negative characteristics of instructors as a function of attractive evaluation method. American Educational Research Journal, 17, 238-245.

Lovell, G. D., & Haner, C. F. (1955). Forced-choice scales applied to college faculty ratings. Educational and Psychological Measurement, 15, 291-304.

Marsh, H. W., Overall, J. U., & Kesler, S. P. (1979). Class size, students' evaluations, and instructional effectiveness. American Educational Research Journal, 16, 57-69.

Marsh, H. W., Overall, J. U., & Thomas, C. S. (1976). The relationship between students' evaluations of instruction and expected grades. (ERIC Document Reproduction Service No. ED 126 140)

McEachi, W. J. (1986). Teaching tips: A guidebook for the beginning college teacher. Toronto: D. C. Heath and Company.

Riley, J. W., Ryan, B. F., & Lifschitz, M. (1950). The student looks at his teacher. New Brunswick, NJ: Rutgers University Press.

Schlenker, D. E., & McKinnon, N. C. (1994). Assessing faculty performance using the student evaluation of instruction. Atlantic Baptist College, New Brunswick, Canada. (ERIC Document Reproduction Service No. ED 371 667)

Table 1

The Results of ANOVA for Course Level by Four Evaluation Dimensions

Dimension	Course Level	Number	Mean	SD	F	Post hoc ^a
Preparation/ Planning	1000	96	3.864	.347	7.919***	4>1
	2000	142	3.951	.408		4>2
	3000	119	3.967	.412		4>3
	4000	83	4.146	.401		
Material/ Content	1000	96	3.967	.298	10.398***	4>1
	2000	142	4.100	.390		4>2
	3000	119	4.100	.389		4>3
	4000	83	4.287	.448		
Method/ Skills	1000	96	3.715	.378	12.357***	4>2
	2000	142	3.852	.438		4>3>1
	3000	119	3.888	.450		
	4000	83	4.106	.466		
Assignment/ Examination	1000	96	3.822	.315	6.266***	4>1
	2000	142	3.905	.352		4>2
	3000	119	3.920	.402		
	4000	83	4.058	.398		

Note. ^a post hoc comparison. The number under the post hoc column indicates the course level, 1:1000, 2: 2000, 3:3000, and 4:4000.

*** $p < .0001$

Table 2

The Results of ANOVA for Class Size by Four Evaluation Dimensions

Dimension	Course Size	Number	Mean	SD	<i>F</i>	<i>p</i>
Preparation/ Planning	<10	72	3.994	.344	.667	.573
	11-20	140	3.985	.434		
	21-30	147	3.936	.445		
	31+	81	4.002	.318		
Material/ Content	<10	72	4.123	.435	.212	.888
	11-20	140	4.108	.450		
	21-30	147	4.085	.378		
	31+	81	4.116	.277		
Method/ Skills	<10	72	3.946	.490	1.520	.209
	11-20	140	3.916	.500		
	21-30	147	3.830	.438		
	31+	81	3.850	.326		
Assignment/ Examination	<10	72	4.000	.367	2.275	.079
	11-20	140	3.946	.425		
	21-30	147	3.879	.369		
	31+	81	3.877	.271		

Table 3

The Results of t Test for Instructor Gender by Four Evaluation Dimensions

Dimension	Instructor	Number	Mean	SD	<i>t</i>	<i>p</i>
Preparation/ Planning	Male	287	3.987	.428	.978	.329
	Female	153	3.947	.356		
Material/ Content	Male	287	4.133	.407	2.135*	.033
	Female	153	4.049	.369		
Method/ Skills	Male	287	3.829	.470	1.723	.086
	Female	153	3.907	.410		
Assignment/ Examination	Male	287	3.936	.389	1.215	.225
	Female	153	3.890	.344		

Note. * $p < .05$